

# The bipenalty method for arbitrary multipoint constraints

Jack Hetherington<sup>1,\*,\dagger</sup>, Antonio Rodríguez-Ferran<sup>2</sup> and Harm Askes<sup>1</sup>

<sup>1</sup>*Department of Civil and Structural Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK*

<sup>2</sup>*Laboratori de Càlcul Numèric, Universitat Politècnica de Catalunya, Jordi Girona 1–3, 08034 Barcelona, Spain*

## SUMMARY

In finite element (FE) analysis, traditional penalty methods impose constraints by adding virtual stiffness to the FE system. In dynamics, this can decrease the critical time step of the system when conditionally stable time integration schemes are used by introducing spurious modes with high eigenfrequencies. Recent studies have shown that using mass penalties alongside traditional stiffness penalties can mitigate this effect for systems with a one single-point constraint. In the present work, we extend this finding to include systems with an arbitrary set of multipoint constraints. By analysing the generalised eigenvalue problem, we show that the values of spurious eigenfrequencies may be controlled by the choice of stiffness and mass penalty parameters. The method is demonstrated using numerical examples, including a one-dimensional contact-impact formulation and a two-dimensional crack propagation analysis. The results show that constraint imposition using the bipenalty method can be employed such that the critical time step of an analysis is unaffected, whereas also displaying superiority over the mass penalty method in terms of accuracy and versatility.

KEY WORDS: solids; finite element methods; penalty methods; stability; time integration, explicit; eigenvalue analysis

## 1. INTRODUCTION

The penalty method is a common technique for imposing constraints in numerical analysis. In the field of finite element (FE) analysis, constraints enforce relationships between DOFs. A simple example of a constraint is the imposition of Dirichlet boundary conditions. For structural FE analysis, this will usually consist of setting the displacement of certain DOF to some known value (most commonly zero). This kind of constraint is referred to as an *absolute* or *single-point* constraint because the constraint equation is written in terms of a single DOF. However, these constraints are simple to impose by transformation of the system equations [1]. Enforcing constraints that describe relationships between more than one DOFs (referred to here as *relative* or *multipoint* constraints) using transformation requires a greater level of manipulation, and for more complex non-linear constraints, the method becomes impractical. Owing in part to their greater level of versatility, penalty methods have become a commonly chosen alternative. Examples include node-to-node tyings [2], interface elements/cohesive surfaces [3, 4] and contact conditions [5, 6]. Despite the fact that penalty methods only approximately impose the given constraints, they are often favoured over the Lagrange multiplier method because they are straightforward to implement, and they do not introduce any extra solution variables. In addition, the classical Lagrange multiplier method is known to be incompatible with certain explicit schemes, including the widely used central difference method (CDM) [7].

---

\*Correspondence to: Jack Hetherington, Department of Civil and Structural Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK.

<sup>\dagger</sup>E-mail: cip09jeh@sheffield.ac.uk

The most popular form of the penalty method, referred to here as the *stiffness penalty method*, imposes an approximate displacement constraint by adding large values into the stiffness matrix of an FE system. In physical terms, the penalties may be thought of as virtual springs of large stiffness. In transient analysis, acceleration constraints may also be imposed using the *mass penalty method*, an analogous technique that adds values to the mass matrix of the system [8]. The physical analogy for this type of constraint is the inerter [9]. Although the implementation of these two methods is very similar, they have very different effects on the eigenvalues of the FE system. It is well known that stiffness penalties introduce spurious modes with very high eigenfrequencies, increasing monotonically as the penalty parameters are increased [10]. On the other hand, mass penalties introduce spurious eigenmodes with very low eigenfrequencies (tending to zero) [11]. The so-called *bipenalty method* utilises both kinds of penalties simultaneously in order to control the eigenfrequencies of these spurious modes. For conditionally stable time integration schemes, this is useful for ensuring that the critical time step of the analysis (which is dependent on the maximum eigenfrequency of the system [12]) is not affected by the imposition of constraints while still taking advantage of established stiffness penalty techniques.

The present investigation builds upon earlier work, which was limited to the application of single-point constraints [13, 14], focusing instead on more general applications, such as interface elements, contact elements and cohesive surfaces. An alternative treatment of the bipenalty method, focussed mainly (although not exclusively) on frequency domain analysis, can also be found in Reference [15], whereas recent work by Ilanko and Monterrubio [16] highlights some of the advantages of the bipenalty method for computing the eigenvalues of constrained systems.

For contact–impact problems, the bipenalty method has been utilised even for implicit, unconditionally stable time integration schemes. The use of penalty methods that enforce displacement constraints only are known to lead to oscillations in the contact forces [5], the origins of which can be ‘traced in part to the lack of satisfaction of the constraint in the velocities’ [17, p. 283]. The bipenalty method has therefore been used before in contact–impact formulations [17, 18] to improve the behaviour of contact algorithms.

It is the goal of this work to examine the bipenalty method for use in a wide range of transient FE analyses. In particular, we focus on the use of the method with the explicit, conditionally stable CDM. In Section 2, we present a versatile formulation for imposing an arbitrary set of constraints. Through eigenvalue analysis, we predict the effects that bipenalisation has on the associated eigenproblem, and suggest guidelines for choosing a suitable ratio of stiffness and mass penalty parameters that will ensure time step stability for conditionally stable schemes such as the CDM. Section 3 provides numerical examples to demonstrate the performance of the bipenalty method in terms of both accuracy and stability in the context of a two-dimensional (2D) interface element formulation.

## 2. BIPENALTY FORMULATION FOR ARBITRARY CONSTRAINTS

In the present work, we consider the FE discretisation of the equations of elastodynamics, represented by the system of  $n$  equations

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f} \quad (1)$$

where  $\mathbf{M}$  is the mass matrix,  $\mathbf{K}$  is the stiffness matrix,  $\mathbf{u}$  is a vector of displacements,  $\mathbf{f}$  a vector of external forces, and dot notation is used to indicate derivatives with respect to time.

### 2.1. Bipenalty matrix formulation

A stiffness tying can be formulated by considering the potential energy,  $\mathcal{U}$ , of the system. Consider a system with  $n$  DOFs and  $k$  linear displacement constraints (which may be absolute or relative in nature) written in the form

$$\mathbf{h} = \mathbf{C}\mathbf{u} - \mathbf{q} \quad (2)$$

where  $\mathbf{C}$  is a matrix of dimension  $k \times n$ , which describes the coupling between DOF, and  $\mathbf{q}$  is a list of prescribed displacement values. Then,  $\mathbf{h} = \mathbf{0}$  implies exact satisfaction of all displacement constraints. We assume here that there are no duplicated constraints, so that the  $k$  rows of  $\mathbf{C}$  are linearly independent.

The standard penalty method adds an extra term, the *penalty function*, to the potential energy of the system, so that

$$\mathcal{U} = \frac{1}{2} \mathbf{u}^T \mathbf{K} \mathbf{u} - \mathbf{u}^T \mathbf{f} + \frac{1}{2} \mathbf{h}^T \mathbf{P}_s \mathbf{h} \quad (3)$$

where  $\mathbf{P}_s$  is a diagonal matrix of stiffness-type penalty parameters, which we will denote by  $\alpha_{s,j}$ , where  $j$  is the number of the constraint equation (i.e. the row and column of the parameter in  $\mathbf{P}_s$ ). The parameters have the same units as their corresponding entry in the matrix  $\mathbf{K}$ , in this case N/m.

In order to introduce the inertia penalty, we also consider constraints on the *rate* of displacement, written as

$$\dot{\mathbf{h}} = \mathbf{C} \dot{\mathbf{u}} - \dot{\mathbf{q}} \quad (4)$$

where it is assumed that  $\mathbf{C}$  is constant. As before,  $\dot{\mathbf{h}} = \mathbf{0}$  implies exact satisfaction of the rate constraints.

Considering now the kinetic energy  $\mathcal{T}$  of the system, with these constraints in place, we have

$$\mathcal{T} = \frac{1}{2} \dot{\mathbf{u}}^T \mathbf{M} \dot{\mathbf{u}} + \frac{1}{2} \dot{\mathbf{h}}^T \mathbf{P}_m \dot{\mathbf{h}} \quad (5)$$

where  $\mathbf{P}_m$  is a diagonal matrix of inertia-type penalty parameters  $\alpha_{m,j}$ , which have units of kg for structural systems. The full dynamic equilibrium equations then follow from

$$\frac{d}{dt} \frac{\partial \mathcal{T}}{\partial \dot{\mathbf{u}}^T} + \frac{\partial \mathcal{U}}{\partial \mathbf{u}^T} = [\mathbf{M} + \mathbf{M}^P] \ddot{\mathbf{u}} + [\mathbf{K} + \mathbf{K}^P] \mathbf{u} = \mathbf{f} + \mathbf{f}^P \quad (6)$$

$$\text{where } \mathbf{K}^P = \mathbf{C}^T \mathbf{P}_s \mathbf{C} \quad (7)$$

$$\mathbf{M}^P = \mathbf{C}^T \mathbf{P}_m \mathbf{C} \quad (8)$$

$$\mathbf{f}^P = \mathbf{C}^T \mathbf{P}_s \mathbf{q} + \mathbf{C}^T \mathbf{P}_m \ddot{\mathbf{q}} \quad (9)$$

It is worth noting also that the prescribed values  $\mathbf{q}$  as given in (2) are usually constant in time, so that  $\ddot{\mathbf{q}} = \mathbf{0}$ , and the second term on the right-hand side of (9) is not required. For example, to implement a simple tying between DOF  $a$  and  $b$ , written as  $u_a - u_b = 0$ , all elements of  $\mathbf{K}^P$ ,  $\mathbf{M}^P$  and  $\mathbf{f}^P$  are zero, except

$$K_{aa}^P = K_{bb}^P = -K_{ab}^P = -K_{ba}^P = \alpha_s \quad (10)$$

$$M_{aa}^P = M_{bb}^P = -M_{ab}^P = -M_{ba}^P = \alpha_m \quad (11)$$

In general, the matrix we must add to implement a tying with inertial penalties has the same form as the standard stiffness tying, although the penalty parameters are different. We can therefore say that inertia penalties are analogous to the stiffness penalty method; the bipenalty method is the use of both types simultaneously. As with traditional methods, the implementation is simple. The formulation of the constraint equation (2) and the equations of motion (6) describe the technique completely—except for the values of penalty parameters  $\alpha_s$  and  $\alpha_m$ , which are to be selected by the analyst. Because it is the magnitude of these parameters relative to the existing entries in  $\mathbf{K}$  and  $\mathbf{M}$ , which dictates the accuracy of the constraint imposition, *penalty factors* are often used to express the magnitude of the penalty, denoted by

$$p_s = \alpha_{s,j} / F_{s,j} \quad (12)$$

$$p_m = \alpha_{m,j} / F_{m,j} \quad (13)$$

and  $F_{s,j}$  and  $F_{m,j}$  are the diagonal entries from the system matrices  $\mathbf{K}$  and  $\mathbf{M}$  that correspond to the DOF associated with the  $j$ th constraint. For multipoint constraints, the maximum system matrix entry from all corresponding DOF is used.

## 2.2. Eigenvalue analysis of the biperenalised system

The unconstrained, unpenalised eigenvalue problem (UP) for an FE or FE system is given by

$$(\mathbf{K} - \lambda_i \mathbf{M}) \mathbf{u}_i = \mathbf{0} \quad (14)$$

where  $\mathbf{K}$  and  $\mathbf{M}$  are the stiffness and mass matrices, respectively, which are assumed to be symmetric, and the eigenvectors  $\mathbf{u}_i$  and corresponding eigenvalues  $\lambda_i$  form the  $n$  solutions (ordered so that  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \equiv \lambda_{\max}$ ). The biperenalised problem (BP) is given by

$$([\mathbf{K} + \mathbf{K}^P] - \tilde{\lambda}_i [\mathbf{M} + \mathbf{M}^P]) \tilde{\mathbf{u}}_i = \mathbf{0} \quad (15)$$

The  $n$  solutions of the BP are given by  $\tilde{\mathbf{u}}_i$  and  $\tilde{\lambda}_i$  (once again ordered so that  $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \dots \leq \tilde{\lambda}_n \equiv \tilde{\lambda}_{\max}$ ).

Our interest in these eigenproblems is due to the stability conditions for the Newmark family of time integration schemes, from which the CDM can be derived. The critical time step for these schemes is given by

$$\Delta t_{\text{crit}} = \frac{\Omega_{\text{crit}}}{\omega_{\max}} \quad (16)$$

where  $\Omega_{\text{crit}}$  is the critical sampling frequency, and  $\omega_{\max}$  is the maximum eigenfrequency of the system [12]. For the CDM, we have  $\Omega_{\text{crit}} = 2$ , a value that is independent of the problem being considered; therefore, it is the maximum eigenfrequency  $\omega_{\max} = \sqrt{\lambda_{\max}}$ , which controls the value of the critical time step. To ensure time step stability, we must choose a time step such that  $\Delta t \leq \Delta t_{\text{crit}}$ .

We now define a single penalty ratio for all constraint equations,

$$R = \frac{\alpha_{s,j}}{\alpha_{m,j}} \quad \text{for } j = 1 \dots k \quad (17)$$

where  $R$  is a scalar constant with units  $\text{s}^{-2}$ . With the assumption that  $R$  is constant for all  $j$ , the magnitudes of the various penalties may vary as long as the *ratio* of parameters is the same for all constraint equations. Then, we have  $\mathbf{P}_s = R \mathbf{P}_m$  and, from (7) and (8),

$$\mathbf{K}^P = R \mathbf{M}^P \quad (18)$$

In this section, we utilise the Rayleigh quotient  $\rho$ , defined as

$$\rho(\mathbf{v}) = \frac{\mathbf{v}^T \mathbf{K} \mathbf{v}}{\mathbf{v}^T \mathbf{M} \mathbf{v}} \quad (19)$$

where  $\mathbf{v}$  is any non-zero vector [12]. A useful property to note is that the Rayleigh quotient of an eigenvector  $\beta \mathbf{u}_i$  (for any  $\beta \neq 0$ ) is the corresponding eigenvalue, that is

$$\rho(\beta \mathbf{u}_i) = \lambda_i \quad (20)$$

The  $k$  constraints given in Equation (2) may be imposed exactly via direct transformation of the system equations [1], a process that would remove  $k$  DOF from the solution. On the other hand, penalisation does not alter the size of the system, therefore, there remain  $n$  eigensolutions where the exact solution has  $n - k$ . In this section, we aim to examine the nature of these additional  $k$  eigensolutions. Our first observation is that it is not possible for all  $n$  eigenmodes to satisfy the constraints.

### Lemma 1

When applying  $k$  linearly independent constraint equations to a system of size  $n$  using the biperenal method, at most  $n - k$  of the eigenmodes associated with the system satisfy those constraints.

*Proof*

The constraint equations may be written

$$\mathbf{h}_i = \mathbf{C}\tilde{\mathbf{u}}_i \quad (21)$$

where  $\mathbf{C}$  is the constraint matrix (of dimension  $k \times n$ ),  $\tilde{\mathbf{u}}_i$  are the eigenvectors of the constrained system and  $\mathbf{h}_i = \mathbf{0}$  implies the satisfaction of all constraints for eigenmode  $i$ . The rank-nullity theorem states that

$$\text{rank}(\mathbf{C}) + \text{nullity}(\mathbf{C}) = n \quad (22)$$

Because the constraint equations are linearly independent,  $\text{rank}(\mathbf{C}) = k$ . Therefore,

$$\text{nullity}(\mathbf{C}) = n - k \quad (23)$$

meaning that  $\mathbf{h}_i = \mathbf{0}$  is possible for  $n - k$  eigenmodes at most.  $\square$

We can use this information to go further and say that, in fact, *exactly*  $n - k$  eigenmodes satisfy the imposed constraints.

*Lemma 2*

When applying  $k$  linearly independent constraint equations to a system of size  $n$  using the bipenalty method, exactly  $n - k$  of the eigenmodes associated with the system satisfy those constraints, whereas exactly  $k$  of the eigenmodes associated with the system do not, for large  $\alpha_m$ .

*Proof*

First, we note that the eigenvectors  $\tilde{\mathbf{u}}_i$  of a FE system may be scaled such that they are  $\mathbf{M}$ -orthonormal, that is,

$$\tilde{\mathbf{u}}_i^T (\mathbf{M} + \mathbf{M}^P) \tilde{\mathbf{u}}_j = \delta_{ij} \quad (24)$$

where  $\delta_{ij}$  is the Kronecker delta. Considering the case where  $i \neq j$  and taking into account (8), we have

$$\tilde{\mathbf{u}}_i^T \mathbf{M} \tilde{\mathbf{u}}_j + (\mathbf{C}\tilde{\mathbf{u}}_i)^T \mathbf{P}_m (\mathbf{C}\tilde{\mathbf{u}}_j) = 0 \quad (25)$$

We now rewrite the matrix of penalty parameters so that  $\mathbf{P}_m = \alpha_m \mathbf{D}_m$  where  $\mathbf{D}_m$  is a diagonal matrix of dimension  $k \times k$ . Furthermore, we assume that all non-zero entries in  $\mathbf{D}_m$  have the same sign; that is, all entries are positive, or all entries are negative. (This does not constitute a significant loss in generality because negative penalties are rarely used, and when they are, it is usually to complement a separate positive-penalty analysis rather than for individual constraint equations.) Then,

$$\tilde{\mathbf{u}}_i^T \mathbf{M} \tilde{\mathbf{u}}_j + \alpha_m (\mathbf{C}\tilde{\mathbf{u}}_i)^T \mathbf{D}_m (\mathbf{C}\tilde{\mathbf{u}}_j) = 0 \quad (26)$$

from which,

$$\lim_{\alpha_m \rightarrow \infty} \left[ (\mathbf{C}\tilde{\mathbf{u}}_i)^T \mathbf{D}_m (\mathbf{C}\tilde{\mathbf{u}}_j) \right] = 0 \quad (27)$$

The assumption that all entries in  $\mathbf{D}_m$  have the same sign means that there cannot be compensation during the matrix multiplication, and hence, we are left with two possibilities for the vector  $\mathbf{C}\tilde{\mathbf{u}}_i$ , assuming large  $\alpha_m$ :

1.  $\mathbf{C}\tilde{\mathbf{u}}_i = \mathbf{0}$ ,
2.  $\mathbf{C}\tilde{\mathbf{u}}_i$  and  $\mathbf{C}\tilde{\mathbf{u}}_j$  are non-zero and orthogonal.

From Lemma 1, the first case is possible for at most  $n - k$  of the  $n$  eigenmodes. The second case is possible for at most  $k$  of the  $n$  modes because the vector  $\mathbf{C}\tilde{\mathbf{u}}_i$  is of dimension  $k$ . Therefore, following directly from the previously mentioned, we can say that

1.  $\mathbf{C}\tilde{\mathbf{u}}_i = \mathbf{0}$  for  $n - k$  of the  $n$  eigenvectors,
2.  $\mathbf{C}\tilde{\mathbf{u}}_i \neq \mathbf{0}$  for  $k$  of the  $n$  eigenvectors.

□

The constrained modes tend to those of the fully constrained system for large penalty parameters, and therefore, from Rayleigh's theorem of separation, the eigenvalues are bounded by those of the unconstrained system. Hence, it is the spurious modes—those which do not satisfy the constraints—that introduce problematic eigenvalues. Next, we address exactly what happens to those eigenvalues.

#### Theorem

For any system subject to  $k$  bipenalty constraints with large  $\alpha_m$ ,  $k$  of the associated eigenvalues tend to the penalty ratio,  $R$ .

#### Proof

Consider the Rayleigh quotient of the BP with the arbitrary vectors  $\mathbf{v}$  replaced by the  $n$  eigenvectors of the BP,

$$\rho_{\text{BP}}(\tilde{\mathbf{u}}_i) = \frac{\tilde{\mathbf{u}}_i^T \mathbf{K} \tilde{\mathbf{u}}_i + \tilde{\mathbf{u}}_i^T \mathbf{K}^p \tilde{\mathbf{u}}_i}{\tilde{\mathbf{u}}_i^T \mathbf{M} \tilde{\mathbf{u}}_i + \tilde{\mathbf{u}}_i^T \mathbf{M}^p \tilde{\mathbf{u}}_i} = \frac{\tilde{\mathbf{u}}_i^T \mathbf{K} \tilde{\mathbf{u}}_i + R \cdot \tilde{\mathbf{u}}_i^T \mathbf{M}^p \tilde{\mathbf{u}}_i}{\tilde{\mathbf{u}}_i^T \mathbf{M} \tilde{\mathbf{u}}_i + \tilde{\mathbf{u}}_i^T \mathbf{M}^p \tilde{\mathbf{u}}_i} \quad (28)$$

For eigenmodes with  $\tilde{\mathbf{u}}_i^T \mathbf{M}^p \tilde{\mathbf{u}}_i \neq 0$ , the penalty terms dominate, so that in the limit

$$\lim_{\alpha_m \rightarrow \infty} \rho_{\text{BP}}(\tilde{\mathbf{u}}_i) = \frac{R \cdot \tilde{\mathbf{u}}_i^T \mathbf{M}^p \tilde{\mathbf{u}}_i}{\tilde{\mathbf{u}}_i^T \mathbf{M}^p \tilde{\mathbf{u}}_i} = R \quad (29)$$

That is, the eigenvalue corresponding to any eigenvector that gives a non-zero  $\tilde{\mathbf{u}}_i^T \mathbf{M}^p \tilde{\mathbf{u}}_i$  tends to  $R$ , the penalty ratio, for large  $\alpha_m$ . From Lemma 2, exactly  $k$  eigenmodes have  $\tilde{\mathbf{u}}_i^T \mathbf{M}^p \tilde{\mathbf{u}}_i \neq 0$ . Therefore, the  $k$  eigenvalues associated with these modes tend to  $R$  for large  $\alpha_m$ . □

This means that we can directly control the eigenvalues associated with the spurious unconstrained eigenmodes. If our goal is to ensure that  $\tilde{\lambda}_{\max} \leq \lambda_{\max}$  (i.e. the maximum eigenvalue is not increased as a result of constraint imposition) then there exists a *critical penalty ratio* (CPR), which should not be exceeded, defined as

$$R_{\text{crit}} = \lambda_{\max} \quad (30)$$

Hence, to ensure that the critical time step  $\Delta t_{\text{crit}}$  is not affected, we must calculate only the maximum eigenvalue of the *unconstrained* system. This may be estimated numerically or analytically by a variety of means.

Because the largest eigenfrequency of a constrained system is less than or equal to the largest eigenfrequency of the unconstrained system (by Rayleigh's theorem of separation), in many cases choosing a sub-CPR would actually *increase* the critical time step, rather than simply keeping it constant [16]. However, to determine the extent of the increase it would then be necessary to compute the maximum eigenfrequency of the *constrained* system.

### 2.3. Choosing a penalty ratio without computing the maximum eigenvalue

Estimating the maximum eigenvalue of a system is a well-known problem in explicit FE analysis because, from (16), it is required for the selection of a suitable time step. A simple and efficient method of achieving this is to use direct iteration (also known as power iteration), which may be used to compute the maximum eigenvalue without solving the full eigenvalue problem. However,

if we assume that reliable methods are already available for computing a stable time step we can choose a safe penalty ratio directly. Equations (16) and (30) show that by choosing a ratio such that

$$R \leq \frac{4}{\Delta t^2} \quad (31)$$

where  $\Delta t$  is the chosen (sub-critical) time step, we can be confident that the spurious eigenfrequencies introduced by the penalty method will not cause the critical time step to decrease past the time step we have chosen for the analysis. In other words, by choosing a time step, we also effectively choose the maximum allowable eigenfrequency; by selecting a penalty ratio according to (31), all spurious eigenfrequencies are kept below that maximum.

#### 2.4. One-dimensional examples

The one-dimensional (1D) examples presented here utilise two-noded linear elements (bar elements) with lumped mass. The elemental structural stiffness and mass matrices are of the form

$$\mathbf{K}^e = \frac{EA}{h} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \mathbf{M}^e = \frac{\rho Ah}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (32)$$

where  $E$ ,  $A$  and  $\rho$  represent the Young's modulus, cross-sectional area and mass density, respectively, and  $h$  is the length of the element. For 1D bar elements, tying nodes is straightforward because there is only one DOF per node and therefore one constraint equation per tying.

**2.4.1. Node-to-node tyings.** We will first demonstrate the bipenalty method using a FE model of a 1D bar. The bar is 1 m long and is discretised into 100 bar elements, with Young's modulus  $E = 1 \text{ N/m}^2$ , mass density  $\rho = 1 \text{ kg/m}^3$ , and cross-sectional area  $A = 0.01 \text{ m}^2$ . Zero-thickness interface elements are placed between each element (nodes are repeated in order to achieve this). Each interface element enforces equal displacement between the two DOF it connects using a form of the penalty method. The bar is fixed at  $x = 0$  and a traction of  $100 \text{ N/m}^2$  is applied at  $x = 1 \text{ m}$  for the first two time steps only. An unusually small time step of  $\Delta t = 0.01 \cdot \Delta t_{\text{crit}}^e = 10^{-4} \text{ s}$  is chosen for the analysis, where  $\Delta t_{\text{crit}}^e$  is the critical time step for a single bar element; this choice ensures that stable results can be obtained when using stiffness penalties only.

The profile given in Figure 1(a) shows the solution at time  $t = 0.5 \text{ s}$  for a bar without interface elements; this represents the reference solution to which the other analyses should converge towards for large penalty parameters. The remaining plots show similar results for the stiffness, mass and bipenalty methods. Figure 1(b) shows that when using stiffness penalties alone, no instability occurs for  $p_s \leq 10^3$ . This is because the penalty is not high enough to increase the maximum eigenfrequency over the limit set by the time step  $\Delta t$ . We note also that relative displacement between repeated nodes is visible on the plot for  $p_s = 10$ , suggesting an inaccurate solution. Instability occurs as the penalty is increased to  $p_s = 10^5$  despite the small time step.

Figure 1(c) shows the result of using mass penalties exclusively. Although there is convergence towards the exact solution for large penalties, we note that for equal stiffness and mass penalty factors, the *accuracy* of constraint imposition for the mass penalty analyses is clearly inferior. However, no instability is observed in this case.

The bipenalty solutions are shown in Figure 1(d). The accuracy of constraint imposition in each case is comparable with the stiffness penalty method, but, comparable with the mass penalty method, no instability occurs when using large penalty parameters. Clearly, the bipenalty method combines the best aspects of stiffness penalties and mass penalties, while avoiding their respective drawbacks.

**2.4.2. Bar impact problem.** Contact-impact problems represent a complex field of research in the area of numerical analysis. This is due in part to their highly non-linear nature and the need for robust and efficient search and contact detection algorithms. Much of this complexity can be avoided by examining 1D problems (which are commonly used to test new algorithms), and therefore, in the present work, we limit ourselves to the impact of 1D bar elements. The formulation is identical to that used for the 1D interfaces used in the previous section, except that the interface elements are



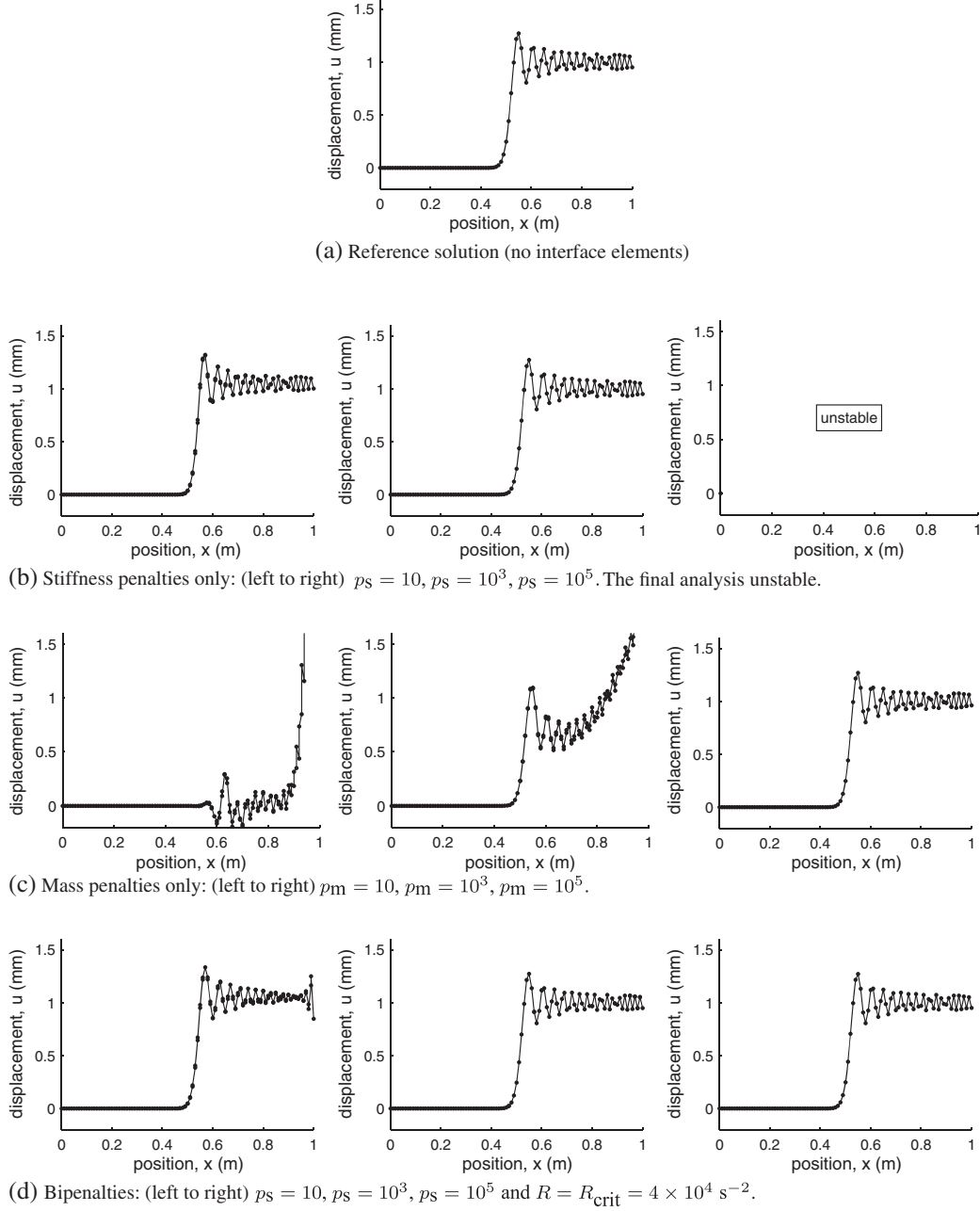


Figure 1. Displacement profiles at time  $t = 0.5 \text{ s}$  for the node-to-node tying experiments. (a) Reference solution (no interface elements); (b) stiffness penalties only: (left to right)  $p_s = 10$ ,  $p_s = 10^3$ ,  $p_s = 10^5$ . The final analysis is unstable; (c) mass penalties only: (left to right)  $p_m = 10$ ,  $p_m = 10^3$ ,  $p_m = 10^5$ ; (d) bipenalties: (left to right)  $p_s = 10$ ,  $p_s = 10^3$ ,  $p_s = 10^5$  and  $R = R_{\text{crit}} = 4 \times 10^4 \text{ s}^{-2}$ .

only active when elements overlap, thereby enforcing equal displacement between nodes while the bars are in contact.

A sketch of the problem we will consider is given in Figure 2 with the relevant data given in Table I. The analysis is run until  $t = 0.6 \text{ s}$  with an initial gap between the bars of zero (i.e. points A and B begin at the same position at  $x = 10 \text{ m}$ ).

An additional complexity when using penalty formulations for contact is introduced by their sensitivity to the choice of penalty parameter. This concept is explored by Huněk [5] whose 1D example



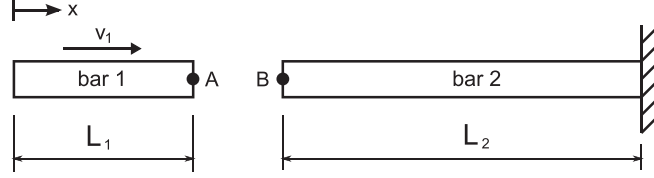


Figure 2. Bar impact problem

Table I. Analysis data for bar impact problem

Property	Bar 1	Bar 2
Length, $L$ (m)	10	20
Young's modulus, $E$ (N/m <sup>2</sup> )	100	100
Density, $\rho$ (kg/m <sup>3</sup> )	0.01	0.01
Cross-sectional area, $A$ (m <sup>2</sup> )	1	1
No. of elements, $N$	50	100
Initial velocity, $v$ (m/s)	0.1	0

we have recreated using the bipenalty method. Essentially, the nature of contact problems introduces an upper bound on the penalty parameter that can be used as high stiffness penalties cause large, spurious oscillations in the contact force. This issue is the focus of the following examples.

It is worth noting that for contact–impact problems, mass penalties cannot be used in isolation. In the example presented here, both bars would have zero acceleration throughout the analysis, and no contact would ever be detected; stiffness penalties must be present in order to detect penetration and apply the initial contact force. For this reason we consider only the stiffness and bipenalty methods in this section.

Figures 3 and 4 show the contact force on the contacting node of bar 1 and the displacement of the two contacting nodes over time. The analytical solution for contact force is 0.05 N for  $t = 0 \dots 0.2$  s and  $t = 0.4 \dots 0.6$  s and zero otherwise. The displacement of the nodes in contact should rise linearly from zero to 10 mm from time  $t = 0 \dots 0.2$  s and then stay constant until falling linearly to zero from time  $t = 0.4 \dots 0.6$  s.

Figure 3 shows the analyses where only stiffness penalties are used. Figure 3(a) gives reasonable results for both contact force and displacement of the contacting nodes, although there are oscillations present in the contact force. As the penalty parameter is increased, the results for both contact force and contact node displacement become less accurate. The initial contact force becomes much larger, introducing significant inertia forces that continue to act even when the contact elements no longer overlap, whereas displacements become erratic and ultimately completely unrealistic. In the extreme case shown in Figure 3(c), we see that the contact force is active for one time step only; this impulse alone is sufficient to prevent the elements from coming into contact with each other for the duration of the analysis. In this case, it is clear that the displacement profile does not converge as the penalty parameter is increased.

In Figure 4, we see the same results with mass penalties included as well as stiffness penalties (with  $\alpha_s = R_{\text{crit}}\alpha_m$ ). In this case, we see that the peak magnitude of the contact force increases as larger penalty parameters are used; however, the displacement profile of the bars are still accurate, even for large penalties.

Finally, Figure 5 shows the error in displacement of a contacting node when compared with the analytical solution. The error does indeed decrease as the penalty parameter is increased, as expected. However, beyond  $\alpha_s = 10^3$  N/m we see that the stiffness penalty solution actually decreases in accuracy as the contact forces become increasingly erratic. The bipenalty method provides a consistently low error because of its additional inertia restraint.

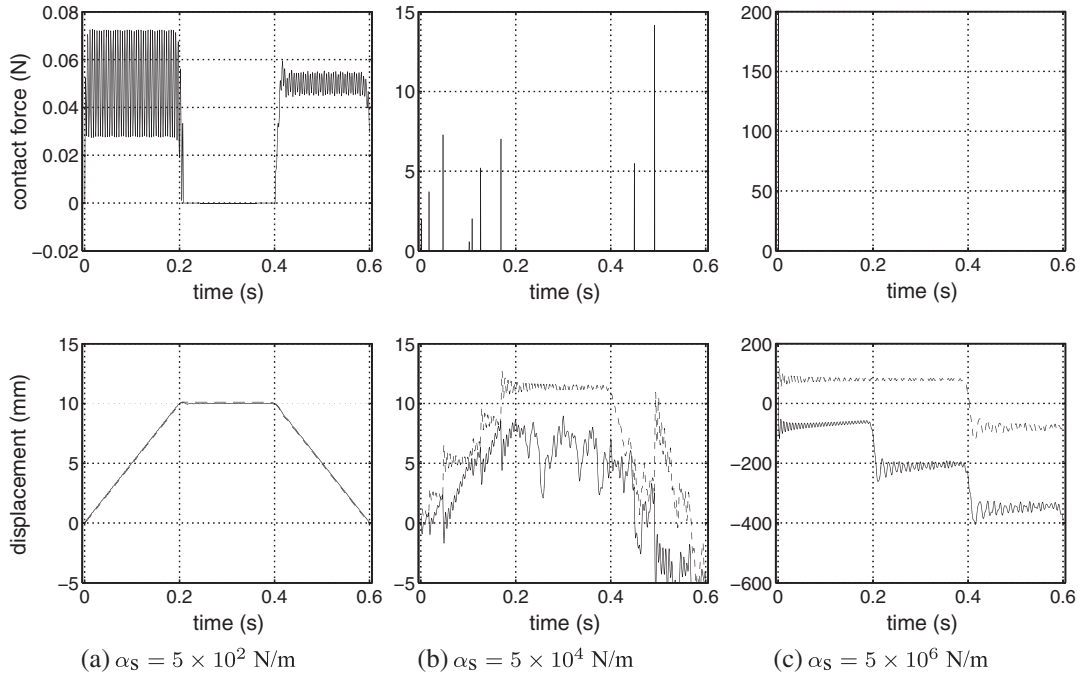


Figure 3. Contact force and displacement of contacting nodes. Stiffness penalties only. (a)  $\alpha_s = 5 \times 10^2$  N/m; (b)  $\alpha_s = 5 \times 10^4$  N/m; and (c)  $\alpha_s = 5 \times 10^6$  N/m.

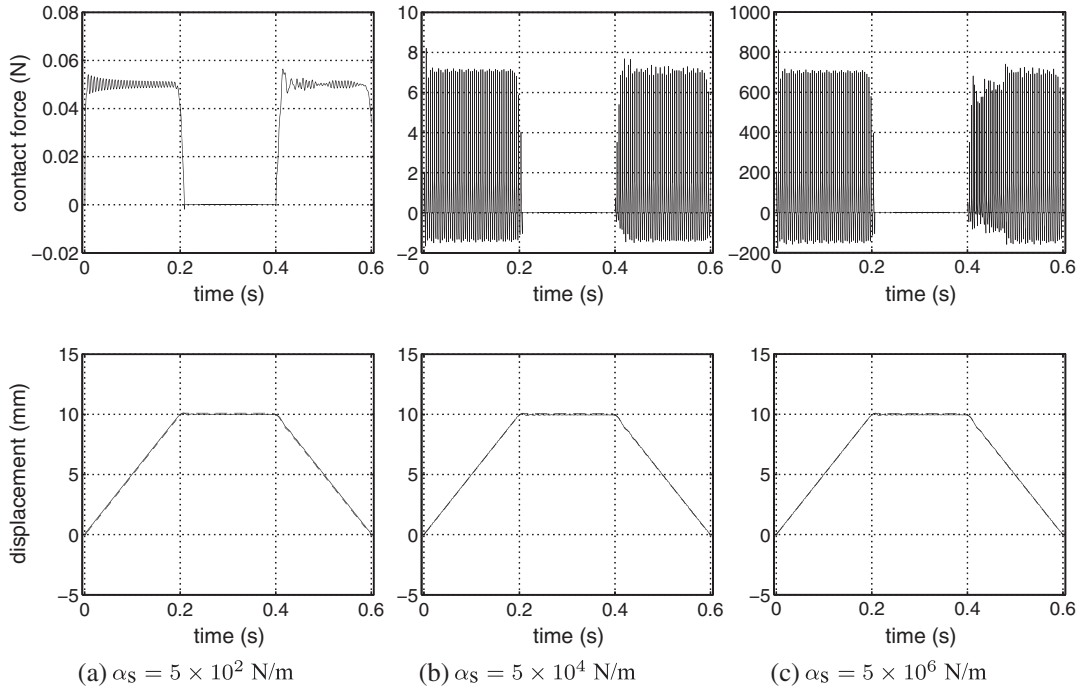


Figure 4. Contact force and displacement of contacting nodes. Bipenalty method with  $R = R_{\text{crit}} = 10^6 \text{ s}^{-2}$ . (a)  $\alpha_s = 5 \times 10^2$  N/m; (b)  $\alpha_s = 5 \times 10^4$  N/m; and (c)  $\alpha_s = 5 \times 10^6$  N/m.

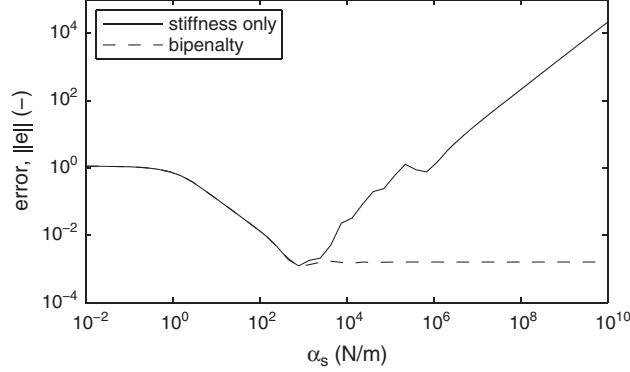


Figure 5.  $L_2$ -norm of displacement error at node A for increasing stiffness penalty.

### 3. A BIPENALTY INTERFACE ELEMENT FORMULATION

There are many ways of modelling crack propagation in FE analysis. For the present work, an interface element formulation is chosen, similar to that introduced by Xu and Needleman [4, 19–21] and developed (for example) by Tjssens [22], as it may easily be modified to incorporate bipenalty methods.

In simple terms, this kind of analysis consists of adding interface elements of zero thickness at all boundaries between FEs lying in areas where cracks may occur. The tractions in the interface elements are then monitored throughout an analysis; when the tractions reach a certain level, the stiffness of the interface elements may be reduced (as in the cohesive surface formulation) or removed completely to create a displacement discontinuity along the path of the crack. The major disadvantage of this approach is that it increases the computational expense of the analysis due to extra DOFs being added (the continuum elements are not directly connected to each other in the global stiffness matrix—only by the cohesion of the interface elements—and hence, nodes must be repeated) and the calculation of tractions within the interface elements themselves. On the other hand, the method is a relatively simple and intuitive extension of standard FE techniques, it allows cracks to form anywhere along the boundaries of the FE mesh, and any number of cracks can form during an analysis.

There are also further and more subtle considerations to be taken into account when using interface elements in this way. In the work of Tjssens [22], for example, it is explained that the constitutive model associated with the interface elements means that some finite displacement must occur as tractions increase, and so we ‘must therefore always check whether the stiffness of the cohesive surfaces is high enough to suppress the elastic deformation of the cohesive surfaces’ and in addition, ‘[t]his is particularly important for applications where cohesive surfaces are immersed through the continuum element mesh’ [22, Appendix A]. Because the path of the crack is generally not known a priori, adding interfaces throughout the mesh is desirable so that cracks can take any path through the material; therefore, giving the interface elements a high stiffness (when using penalty methods, this means high penalty parameters) is of great importance if we wish to obtain an accurate picture of displacements in the continuum. However, interface elements of high stiffness can cause problems in explicit dynamics because of the introduction of spurious eigenfrequencies of large magnitude. Thus, users are faced with a problem: a high stiffness is required in order to achieve the required accuracy, but as a consequence, the critical time step of the analysis is decreased, which may greatly increase computational cost.

The aim of this section is to investigate how the bipenalty method might apply to the use of 2D interface elements in the context of crack propagation analyses. This is achieved by formulating a mass penalty matrix to complement the stiffness penalty matrix of a traditional interface element formulation.

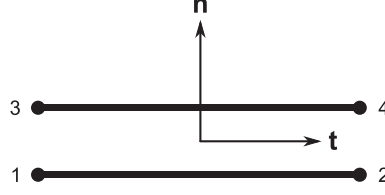


Figure 6. Line interface element, showing node numbering and local coordinate system

### 3.1. Element formulation

The interface element formulation that is the basis for this new penalty-type formulation is based on the work of Schellekens [3, 23]. In this work, a four-noded interface element is derived, which has an initial volume of zero. The stiffness of the element is controlled by user-defined parameters that describe the constitutive behaviour of the element. The goal of the present work is to derive a mass matrix for the interface element that will also limit relative normal and tangential acceleration. This should also enable precise control over the element's maximum eigenvalue. For simplicity, we will be working in 2D and using the 'linear line interface'—an element which is described by normal and tangential tractions and with linear shape functions (see Figure 6).

We now consider this four-noded line interface. Each node has two DOFs, giving an element nodal displacement vector

$$\mathbf{d} = [d_n^1, d_n^2, d_n^3, d_n^4, d_t^1, d_t^2, d_t^3, d_t^4]^T \quad (33)$$

where  $n$  and  $t$  denote the directions normal and tangential to the interface, respectively, and superscripts indicate the node numbers as shown in Figure 6. The continuous displacement field is then

$$\mathbf{u} = [u_n^u, u_n^l, u_t^u, u_t^l]^T \quad (34)$$

where  $u$  and  $l$  denote the upper and lower sides of the interface, respectively. The relationship between nodal and continuous displacement vectors is given by

$$\mathbf{u} = \mathbf{H}\mathbf{d} \quad (35)$$

where  $\mathbf{H}$  contains the interpolation polynomials  $\mathbf{n} = [N_1, N_2]$  and is of the form

$$\mathbf{H} = \begin{bmatrix} \mathbf{n} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{n} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{n} \end{bmatrix} \quad (36)$$

We can find the relative displacements  $\Delta\mathbf{u} = [\Delta u_n, \Delta u_t]^T$  from

$$\Delta\mathbf{u} = \mathbf{L}\mathbf{u} \quad (37)$$

where the operator matrix  $\mathbf{L}$  is defined through

$$\mathbf{L} = \begin{bmatrix} -1 & +1 & 0 & 0 \\ 0 & 0 & -1 & +1 \end{bmatrix} \quad (38)$$

The relationship between nodal displacements and relative displacements can then be derived from Equations (35) and (37) as

$$\Delta\mathbf{u} = \mathbf{LH}\mathbf{d} \quad (39)$$

at which point it is useful to define a matrix  $\mathbf{B}$ , which directly relates relative displacements to nodal displacements

$$\mathbf{B} := \mathbf{LH} = \begin{bmatrix} -\mathbf{n} & \mathbf{n} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{n} & \mathbf{n} \end{bmatrix} \quad (40)$$

For arbitrarily orientated elements, the matrix  $\mathbf{B}$  should be transformed to the local tangential coordinate system of the node set. We now introduce a matrix describing constitutive relations

$$\mathbf{D}_s = \begin{bmatrix} \beta_n & 0 \\ 0 & \beta_t \end{bmatrix} \quad (41)$$

where  $\beta_n$  and  $\beta_t$  describe the stiffness of the interface in the normal and tangential directions. These are our penalty parameters, which should be large in order to keep relative displacements across the interface small. For simplicity, these values are kept constant; for a cohesive surface formulation, the matrix  $\mathbf{D}_s$  could also be written as a function of some damage parameter. The traction versus relative displacement relation is then

$$\mathbf{t} = \mathbf{D}_s \Delta \mathbf{u} \quad (42)$$

and  $\mathbf{t} = [t_n, t_t]^T$  is the traction vector for the element.

The stiffness matrix  $\mathbf{K}$  can now be obtained by minimisation of the total potential energy. The internal work done in the element is

$$\mathcal{U} = \frac{1}{2} \int_S \Delta \mathbf{u}^T \mathbf{t} dS \quad (43)$$

which can be rewritten as

$$\mathcal{U} = \frac{1}{2} \mathbf{d}^T \int_S \mathbf{B}^T \mathbf{D}_s \mathbf{B} dS \mathbf{d} \quad (44)$$

whereas the external work  $\mathcal{W}$  is given by

$$\mathcal{W} = -\mathbf{d}^T \mathbf{f} \quad (45)$$

After variation of the total potential energy  $(\mathcal{U} + \mathcal{W})$ , we find

$$\mathbf{K} \mathbf{d} = \mathbf{f} \quad (46)$$

where the stiffness matrix is given by

$$\mathbf{K} = \int_S \mathbf{B}^T \mathbf{D}_s \mathbf{B} dS \quad (47)$$

If using numerical integration, we can integrate over the isoparametric coordinate  $\xi$ , because for a line interface, the shape functions can be written with one variable only, that is

$$\mathbf{K} = b \int_{-1}^1 \mathbf{B}^T \mathbf{D}_s \mathbf{B} \sqrt{\left(\frac{\partial x}{\partial \xi}\right)^2 + \left(\frac{\partial y}{\partial \xi}\right)^2} d\xi \quad (48)$$

where  $b$  is the width of the interface in the out-of-plane direction.

In order to introduce inertia effects, and obtain a suitable mass matrix for the formulation, we must also consider the kinetic energy of the interface, which is related to velocity. Thus, analogous to the corresponding displacement terms we have

$$\dot{\mathbf{u}} = \mathbf{H} \dot{\mathbf{d}} \quad (49)$$

$$\Delta \dot{\mathbf{u}} = \mathbf{L} \dot{\mathbf{u}} \quad (50)$$

$$\Delta \dot{\mathbf{u}} = \mathbf{B} \dot{\mathbf{d}} \quad (51)$$

We then introduce a momentum vector  $\mathbf{p} = [p_n, p_t]^T$  which is given by

$$\mathbf{p} = \mathbf{D}_m \Delta \dot{\mathbf{u}} \quad (52)$$

where the matrix  $\mathbf{D}_m$  is a matrix containing mass penalties in the normal and tangential directions and is assumed to be a scalar multiple of the constitutive matrix (41) so that  $\mathbf{D}_s = R \mathbf{D}_m$ . The kinetic energy of the interface is then given by

$$\mathcal{T} = \frac{1}{2} \int_S \Delta \dot{\mathbf{u}}^T \mathbf{p} dS \quad (53)$$

which, after invoking (51) and (52) becomes

$$\mathcal{T} = \frac{1}{2} \dot{\mathbf{d}}^T \int_S \mathbf{B}^T \mathbf{D}_m \mathbf{B} dS \dot{\mathbf{d}} \quad (54)$$

The equations of motion then follow from the minimisation of energy

$$\mathbf{M} \ddot{\mathbf{d}} + \mathbf{K} \mathbf{d} = \mathbf{f} \quad (55)$$

where the mass matrix is given by

$$\mathbf{M} = \int_S \mathbf{B}^T \mathbf{D}_m \mathbf{B} dS = \frac{1}{R} \mathbf{K} \quad (56)$$

and hence, as with the penalty matrices derived earlier, the mass matrix is a scalar multiple of the stiffness matrix. This means that the theories of Section 2.2 also apply here, with the assembled interface matrices acting as the penalty matrices  $\mathbf{K}^P$  and  $\mathbf{M}^P$ .

### 3.2. Two-dimensional examples

In the following examples, interface elements are placed between all elements in a certain area by repeating the relevant nodes after meshing. The parameters  $\beta_n$ ,  $\beta_t$  and the penalty ratio  $R$  then allow us to control the level of stiffness and inertial restraint. For simplicity, we have chosen to use a single penalty parameter  $\beta = \beta_n = \beta_t$  so that restraint in normal and tangential directions is equal. Where interfaces are allowed to break, they do so when interface tractions normal to the interface exceed a certain limit, given by  $\sigma_{\max}$ . Only tensile tractions cause cracking, which is appropriate for materials with high compressive strength such as concrete. Interface tractions are calculated by

$$\boldsymbol{\sigma} = \mathbf{D}_s \Delta \mathbf{u} + \mathbf{D}_m \Delta \ddot{\mathbf{u}} \quad (57)$$

where  $\boldsymbol{\sigma} = [\sigma_n, \sigma_t]^T$  are the normal and tangential tractions. An interface is removed when  $\sigma_n > \sigma_{\max}$ , with no cohesion after the peak traction is reached.

**3.2.1. Wave propagation in a halfspace.** In this section, we consider the propagation of stress waves in a block of material of length 2 m and height 1 m. The examples here are fully linear (i.e. no interface failure is considered). A structured mesh of 5000 four-noded square elements is used to mesh the domain. A force is applied to the block from time  $t = 0$  to  $t = 0.1$  s at the centre of the top face in the negative  $y$ -direction. The resulting displacement profile is computed using the CDM until time  $t = 0.8$  s. To assess the accuracy of the penalty formulations, non-breaking interfaces are added between all elements in the right-hand half of the block. A reference solution can be obtained by omitting this step, leaving all elements connected in the normal way.

Figure 7 shows the results of using stiffness penalty interfaces of low stiffness. Figure 7(a) shows, qualitatively, the (symmetrical) Von Mises stress profile obtained when interface elements are absent, whereas Figure 7(b) shows that introducing interface elements causes a significant change in behaviour on the right-hand side of the block. The most obvious effect is the lower velocity of the stress wave, which is a result of decreased stiffness. Figure 7(c) shows the magnitude of the error field,

$$e = \sigma_v - \sigma_v^{\text{ref}} \quad (58)$$

where  $\sigma_v$  and  $\sigma_v^{\text{ref}}$  contain the element-wise Von Mises stress values for the test solution and reference solution, respectively. A global quantity that gives a measure of error across the field is provided by the  $L_2$ -norm of  $e$ , denoted here by  $\|e\|$ .

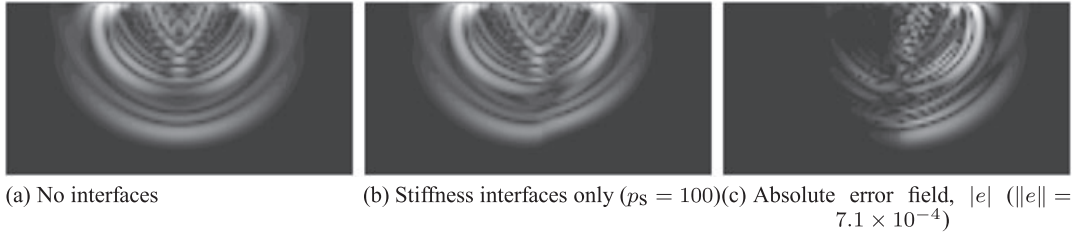


Figure 7. Von Mises stress profiles (and error field) for an experiment using a low stiffness penalty. (a) No interfaces; (b) stiffness interfaces only ( $p_s = 100$ ); and (c) absolute error field,  $|e|$  ( $\|e\| = 7.1 \times 10^{-4}$ ).

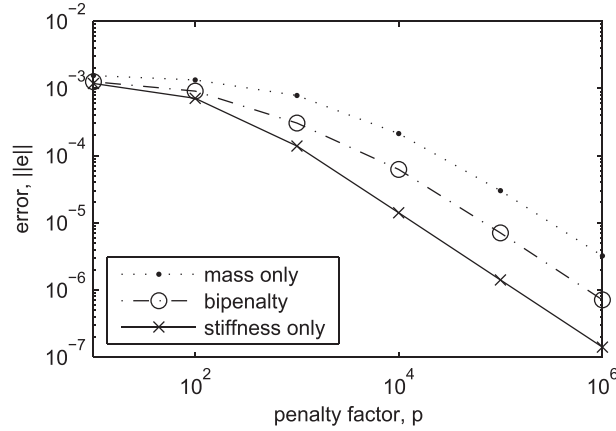


Figure 8. Constraint imposition error.

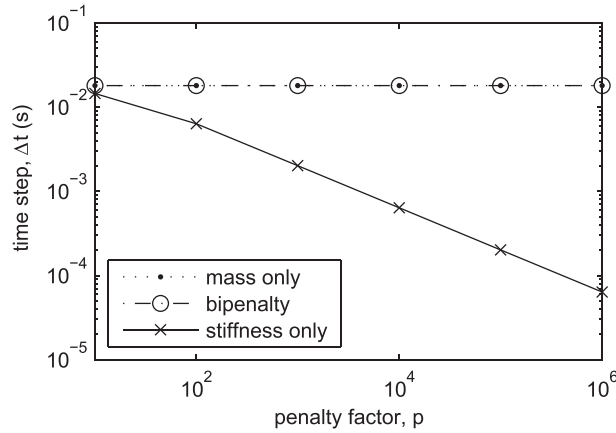


Figure 9. Stable time step.

The norm of the error profile  $\|e\|$  and the time step  $\Delta t$  used to achieve stability are plotted in Figures 8 and 9, respectively, for a number of similar tests. The penalty factor,  $p$ , refers to the mass penalty factor  $p_m$  when the mass penalty method is used and the stiffness penalty factor  $p_s$  otherwise. When the bipenalty method is used,  $p_s = p_m$ . For all forms of the penalty method, the accuracy of the stress profile is improved as the penalty factor is increased. Also, for similar penalty factors, the stiffness and bipenalty methods give more accurate results than the mass penalty method. However, only the stiffness penalty method requires an alteration of the time step. Indeed, to achieve an error norm lower than  $10^{-5}$ , the time step must be reduced by two orders of magnitude.



Thus, the bipenalty method appears to be a suitable alternative to standard penalty methods when high accuracy is required, but using small time steps is undesirable.

**3.2.2. Crack propagation in a single-edge notched beam.** We now consider a single-edge notched concrete beam as described in Figure 10. A crack propagation analysis is performed by introducing interface elements, which are removed once a traction value of  $\sigma_{\max} = 3 \text{ kPa}$  is reached. The Young's modulus, mass density and Poisson ratio are taken as 35 GPa,  $1000 \text{ kg/m}^3$  and 0.2, respectively. The applied force,  $P$ , increases linearly from 0 to 100 MN during the analysis. The loading and support platens are 20 mm wide, and the central notch has a length and width of 20 and 5 mm, respectively.

For efficiency, interface elements are not placed throughout the entire mesh, but only in the central area of the beam between the notch and the main loading area, where the main crack path is expected to develop (this area is referred to here as the process window). Suitable bounds for this area can be established by first experimenting with a coarse FE mesh and/or analysis of stress profiles.

The crack paths for three analyses are shown in Figure 11. For this example, we have used the same time step for all analyses, calculated by multiplying the maximum elemental critical time step by a safety factor of 0.9, giving  $\Delta t \approx 31 \text{ } \mu\text{s}$ . (For crack propagation examples, we must use elemental eigenfrequencies in the time step calculation rather than the eigenfrequencies of assembled element systems because the system eigenfrequencies may increase as cracks form.)

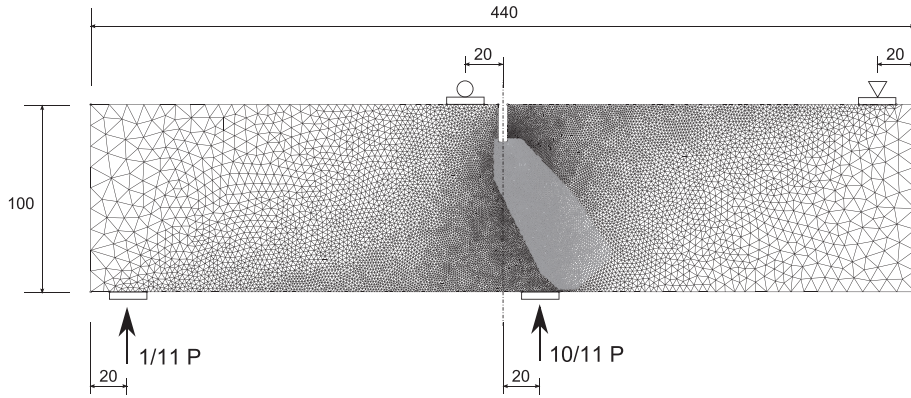
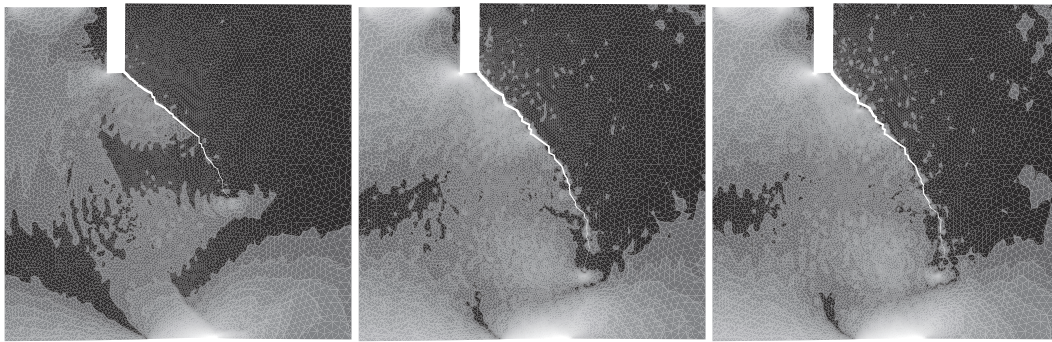


Figure 10. Mesh used for the single-edge notched beam, with support and loading conditions shown. The grey area in the centre shows the extent of the process window.



(a) Stiffness penalty method,  $p_s = 0.425$  (maximum stable penalty) (b) Bipenalty method,  $p_s = p_m = 10^6$  (c) Mass penalty method,  $p_m = 10^6$

Figure 11. Von Mises stress profiles at time  $t = 0.2 \text{ s}$ . Displacements have been magnified by a factor of 100. (a) Stiffness penalty method,  $p_s = 0.425$  (maximum stable penalty); (b) bipenalty method,  $p_s = p_m = 10^6$ ; (c) mass penalty method,  $p_m = 10^6$ .

This means that the maximum stiffness penalty factor (as employed in Figure 11(a)) is rather low, causing clear discontinuities in the stress profile around the edges of the process zone. For the mass and bipenalty methods, much larger penalty parameters may be used, and these analyses both predict a longer, more advanced crack path.

Although this model does not take into account damage/cohesion, the predicted crack path is similar to quasi-static experimental results found elsewhere in the literature [24, 25].

#### 4. CONCLUSIONS

The bipenalty formulation presented in this paper is a straightforward combination of stiffness and mass penalty methods. For 1D tyings and 2D interface elements, the bipenalty method offers constraint imposition errors comparable to those found when using standard stiffness penalty methods, the advantage being that spurious eigenvalues may be controlled using the ratio of penalty parameters so that stability can be ensured without decreasing time step size. For 1D contact problems, the introduction of mass penalties alongside stiffness penalties can increase the accuracy of displacement profiles while again keeping spurious eigenvalues low. In this case, it is not possible to use mass penalties alone. For these reasons, bipenalties appear to be a viable and versatile alternative to traditional penalty methods.

So far, it has been argued that it is best to choose moderate values for the penalty ratio (i.e. not very large nor very small) because this allows to control accuracy as well as wave propagation speeds and thus, in turn, the critical time step. However, it is noted that using bipenalties does introduce a finite eigenfrequency that could lead to resonance. That is, enforcement of the constraints with the bipenalty method could fail if a periodic force with frequency  $\omega \equiv \sqrt{R}$  is applied. This issue would appear to be more relevant in vibration analysis than in wave propagation analysis, but nevertheless could warrant further study.

Another issue that has not been addressed in this paper is the effect the penalty ratio has on the accuracy of solutions. The CPR  $R_{\text{crit}}$  gives an upper bound on the penalty ratio required for stability; it is not necessarily optimal. Further analysis is required to determine to what extent the ratio of penalties affects the accuracy of constraint imposition in differing circumstances.

#### ACKNOWLEDGEMENT

The authors gratefully acknowledge financial support from the Royal Society through International Joint Project JP090621, ‘Bipenalty method for finite elements and explicit time integration’.

#### REFERENCES

1. Cook RD, Malkus DS, Plesha ME. *Concepts and Applications of Finite Element Analysis*, 3rd ed. John Wiley & Sons: New York, 1989.
2. Askes H, Piercy S, Ilanko S. Tyings in linear systems of equations modelled with positive and negative penalty functions. *Communications in Numerical Methods in Engineering* 2007; **24**(11):1163–1169.
3. Schellekens JCJ, De Borst R. On the numerical integration of interface elements. *International Journal for Numerical Methods in Engineering* 1993; **36**:43–66.
4. Xu XP, Needleman A. Numerical simulations of dynamic crack growth along an interface. *International Journal of Fracture* 1996; **74**(4):289–324.
5. Huněk I. On a penalty formulation for contact–impact problems. *Computers & Structures* 1993; **48**(2):193–203.
6. Belytschko T, Neal MO. Contact–impact by the pinball algorithm with penalty and Lagrangian methods. *International Journal for Numerical Methods in Engineering* 1991; **31**(3):547–572.
7. Carpenter NJ, Taylor RL, Katona MG. Lagrange constraints for transient finite element surface contact. *International Journal for Numerical Methods in Engineering* 1991; **32**:103–128.
8. Hetherington J, Askes H. Penalty methods for time domain computational dynamics based on positive and negative inertia. *Computers & Structures* 2009; **87**(23–24):1474–1482.
9. Smith MC. Synthesis of mechanical networks: the inerter. *IEEE Transactions on Automatic Control* 2002; **47**(10):1648–1662.
10. Ilanko S. Existence of natural frequencies of systems with artificial restraints and their convergence in asymptotic modelling. *Journal of Sound and Vibration* 2002; **255**(5):883–898.

11. Ilanko S. Introducing the use of positive and negative inertial functions in asymptotic modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2005; **461**:2545–2562.
12. Hughes TJR. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Prentice-Hall: Englewood Cliffs, NJ, 1987.
13. Askes H, Carames-Saddler M, Rodriguez-Ferran A. Bipenalty method for time domain computational dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2010; **466**:1389–1408.
14. Hetherington J, Rodríguez-Ferran A, Askes H. A new bipenalty formulation for ensuring time step stability in computational dynamics. *International Journal for Numerical Methods in Engineering* 2012; **90**(3):269–286.
15. Paraskevopoulos EA, Panagiotopoulos CG, Manolis GD. Imposition of time-dependent boundary conditions in FEM formulations for elastodynamics: critical assessment of penalty-type methods. *Computational Mechanics* 2010; **45**(2–3):157–166.
16. Ilanko S, Monterrubio LE. Bipenalty method from a frequency domain perspective. *International Journal for Numerical Methods in Engineering* 2012; **90**(10):1278–1291.
17. Armero F, Petocz E. Formulation and analysis of conserving algorithms for frictionless dynamic contact/impact problems. *Computer Methods in Applied Mechanics and Engineering* 1998; **158**(3–4):269–300.
18. Asano N. A virtual work principle using penalty function method for impact contact problems of two bodies. *Bulletin of the JSME* 1986; **29**(249):731–736.
19. Xu XP, Needleman A. Void nucleation by inclusion debonding in a crystal matrix. *Modelling and Simulation in Materials Science and Engineering* 1993; **1**(2):111–132.
20. Xu XP, Needleman A. Numerical simulations of fast crack growth in brittle solids. *Journal of the Mechanics and Physics of Solids* 1994; **42**(9):1397–1434.
21. Xu XP, Needleman A. Numerical simulations of dynamic interfacial crack growth allowing for crack growth away from the bond line. *International Journal of Fracture* 1995; **74**(3):253–275.
22. Tijssens MGA. On the cohesive surface methodology for fracture of brittle heterogeneous solids. *Ph.D. Thesis*, Delft University of Technology, 2000.
23. Schellekens JCJ. Computational strategies for composite structures. *Ph.D. Thesis*, Delft University of Technology, Delft, 1992.
24. Schlangen E. Experimental and numerical analysis of fracture processes in concrete. *Ph.D. Thesis*, Delft University of Technology, 1993.
25. Simone A. Continuous–discontinuous modelling of failure. *Ph.D. Thesis*, Delft University of Technology, 2003.